Independence

CSE 312 Spring 24
Lecture 8

Announcements

HW2 due tonight

HW3 out this evening.

HW3 includes a programming problem – using Bayes rule to do some machine learning – detecting whether emails are spam or “ham” (legitimate emails).

Longer than the programming on HW1 – please get started early!

Extra resources will be available!

Today

Outline:

Independence

Chain Rule

Conditional Independence

Definition of Independence

We’ve calculated conditional probabilities.

Sometimes conditioning – getting some partial information about the outcome – doesn’t actually change the probability.

We already saw an example like this…

Conditioning Practice

Red die 6 conditioned on sum 7 1/6

Red die 6 conditioned on sum 9 1/4

Sum 7 conditioned on red die 6 1/6

Red die 6 has probability 1/6 before or after conditioning on sum 7.

	D2=1	D2=2	D2=3	D2=4	D2=5	D2=6
D1=1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1.6)
D1=2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
D1=3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
D1=4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
D1=5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
D1=6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Independence

Independence

Two events A, B are independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

You'll sometimes see this called "statistical independence" to emphasize that we're talking about probabilities (not, say, physical interactions).

If A, B both have non-zero probability then

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \Leftrightarrow \mathbb{P}(A|B) = \mathbb{P}(A) \Leftrightarrow \mathbb{P}(B|A) = \mathbb{P}(B)$$

Examples

We flip a fair coin three times. Each flip is independent. (both in the statistical independence sense and in the "doesn't affect the next one" sense).

Is $E = \{HHH\}$ independent of $F =$"at most two heads"?

Are $A =$"the first flip is heads" and $B =$"the second flip is tails" independent?

Examples

Is $E = \{HHH\}$ independent of $F =$"at most two heads"?

$\mathbb{P}(E \cap F) = 0$ (can't have all three heads and at most two heads).

$\mathbb{P}(E) = 1/8$, $\mathbb{P}(F) = 7/8$, $\mathbb{P}(E \cap F) \neq \mathbb{P}(E)\mathbb{P}(F)$.

Are $A =$"the first flip is heads" and $B =$"the second flip is tails" independent?

$\mathbb{P}(A \cap B) = 2/8$ (uniform measure, and two of eight outcomes meet both A and B.

$\mathbb{P}(A) = 1/2$, $\mathbb{P}(B) = 1/2$ $\frac{2}{8} = \frac{1}{2} \cdot \frac{1}{2}$. These are independent!

Hey Wait

I said “the flips are independent” why aren’t E, F independent?

“the flips are independent” means events like <the first flip is blah>” is independent of events like <the second flip is blah>

But if you have an event that involves both flip one and two that might not be independent of an event involving flip one or two.

Mutual Exclusion and independence

Two of these statements are true, one is false. Explain to each other which ones are true, and find a counter-example to the false one.

1. If A, B both have nonzero probability and they are mutually exclusive, then they cannot be independent.

2. If A has zero probability, then A, B are independent (for any B).

3. If two events are independent, then at least one has nonzero probability.

Fill out the poll everywhere so Robbie knows how long to explain
Go to pollev.com/robbie

Mutual Exclusion and independence

Two of these statements are true, one is false. Explain to each other which ones are true, and find a counter-example to the false one.

1. If A, B both have nonzero probability and they are mutually exclusive, then they cannot be independent.

2. If A has zero probability, then A, B are independent (for any B).

3. If two events are independent, then at least one has nonzero probability.

Chain Rule

Chain Rule

We defined conditional probability as: $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$

Which means $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$

Chain Rule

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n)$$
$$= \mathbb{P}(A_n | A_1 \cap \cdots \cap A_{n-1}) \cdot \mathbb{P}(A_{n-1} | A_1 \cap \cdots \cap A_{n-2}) \cdots \mathbb{P}(A_2 | A_1) \cdot \mathbb{P}(A_1)$$

Chain Rule Example

Shuffle a standard deck of 52 cards (so every ordering is equally likely).

Let A be the event "The top card is a K♦"

Let B be the event "the second card is a J ♠

Let C be the event "the third card is a 5 ♠

What is $\mathbb{P}(A \cap B \cap C)$?

Use the chain rule!

$$\mathbb{P}(A) \cdot \mathbb{P}(B|A) \cdot \mathbb{P}(C|A \cap B)$$

$$\frac{1}{52} \cdot \frac{1}{51} \cdot \frac{1}{50}$$

Conditional Independence

Conditional Independence

We say A and B are conditionally independent on C if

$\mathbb{P}(A \cap B|C) = \mathbb{P}(A|C) \cdot \mathbb{P}(B|C)$

i.e. if you condition on C, they are independent.

Conditional Independence Example

You have two coins. Coin A is fair, coin B comes up heads with probability 0.85.

You will roll a (fair) die, if the result is odd flip coin A twice (independently); if the result is even flip coin B twice (independently)

Let C_1 be the event "the first flip is heads", C_2 be the event "the second flip is heads", O be the event "the die was odd"

Are C_1 and C_2 independent? Are they independent conditioned on O?

(Unconditioned) Independence

$\mathbb{P}(C_1) = \mathbb{P}(O)\mathbb{P}(C_1|O) + \mathbb{P}(\bar{O})\mathbb{P}(C_1|\bar{O})$

$= \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot 0.85 = .675$

$\mathbb{P}(C_2) = .675$ (the same formula works)

$\mathbb{P}(C_1)\mathbb{P}(C_2) = .675^2 = .455625$

$\mathbb{P}(C_1 \cap C_2) = \mathbb{P}(O)\mathbb{P}(C_1 \cap C_2|O) + \mathbb{P}(\bar{O})\mathbb{P}(C_1 \cap C_2|\bar{O})$

$= \frac{1}{2} \cdot \frac{1}{4} + \frac{1}{2} \cdot .85^2 = .48625$

Those aren't the same! They're not independent!

Intuition: seeing a head gives you information – information that it's more likely you got the biased coin and so the next head is more likely.

Conditional Independence

$\mathbb{P}(C_1|O) = 1/2$

$\mathbb{P}(C_2|O) = 1/2$

$\mathbb{P}(C_1 \cap C_2|O) = \frac{1}{2} \cdot \frac{1}{2} = 1/4$

$\mathbb{P}(C_1|O)\mathbb{P}(C_2|O) = \mathbb{P}(C_1 \cap C_2|O)$

Yes! C_1 and C_2 are conditionally independent, conditioned on O.

Takeaway

Read a problem carefully – when we say “these steps are independent of each other” about some part of a sequential process, it’s usually “conditioned on all prior steps, these steps are conditionally independent of each other.”

Those conditional steps are usually dependent (without conditioning) because they might give you information about which branch you took.

Setting the stage: Random Variables

Implicitly defining Ω

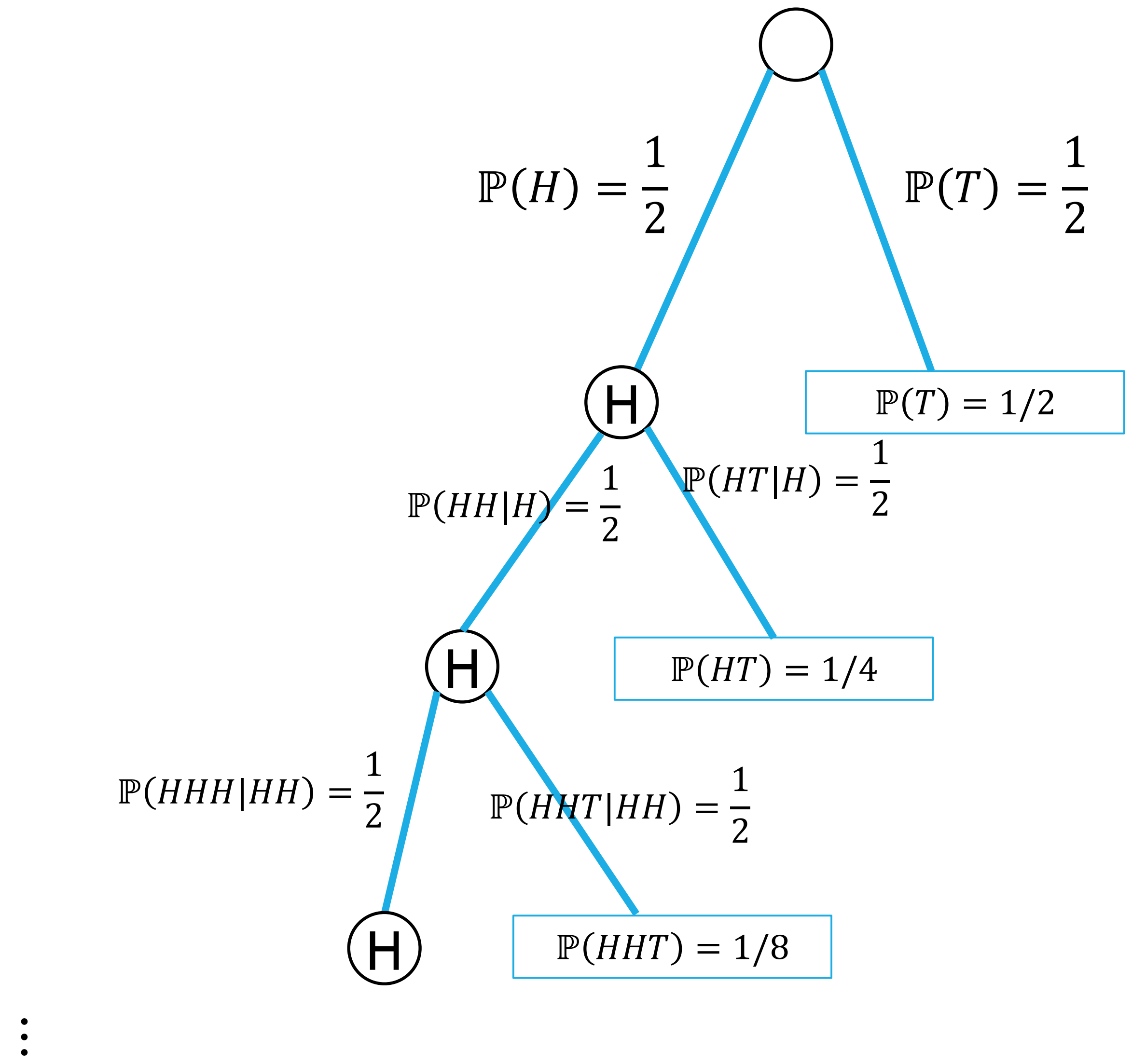
We've often skipped an explicit definition of Ω.

Often $|\Omega|$ is infinite, so we really couldn't write it out (even in principle).

How would that happen?

Flip a fair coin (independently each time) until you see your first tails.
what is the probability that you see at least 3 heads?

An infinite process.



Ω is infinite.

A sequential process is also going to be infinite…

But the tree is "self-similar"

From every node, the children look identical (H with probability ½, continue pattern; T to a leaf with probability ½)

Finding $\mathbb{P}$(at least 3 heads)

Method 1: infinite sum.

Ω includes H^iT for every i. Every such outcome has probability $1/2^{i+1}$

What outcomes are in our event?

$\sum_{i=3}^{\infty} 1/2^{i+1} = \frac{\frac{1}{2^4}}{1-1/2} = \frac{1}{8}$

Infinite geometric series, where common ratio is between -1 and 1 has closed form $\frac{\text{first term}}{1-\text{ratio}}$

Finding $\mathbb{P}$(at least 3 heads)

Method 2:

Calculate the complement

$\mathbb{P}$(at most 2 heads) = $\frac{1}{2}+\frac{1}{4}+\frac{1}{8}$

$\mathbb{P}$(at least 3 heads)= $1-\left(\frac{1}{2}+\frac{1}{4}+\frac{1}{8}\right)=\frac{1}{8}$

Chain Rule

We defined conditional probability as: $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$

Which means $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$

Chain Rule

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n)$$
$$= \mathbb{P}(A_n|A_1 \cap \cdots \cap A_{n-1}) \cdot \mathbb{P}(A_{n-1}|A_1 \cap \cdots \cap A_{n-2}) \cdots \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_1)$$